

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 10:38:31 ; Search time 29 Seconds  
(without alignments)  
85.261 Million cell updates/sec

Title: US-09-551-151a-43

Perfect score: 64

Sequence: 1 SPQGIAGRNFN 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 500 summaries

Database :

1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*  
15: SP viirus: \*  
16: SP bacteriophage: \*  
17: SP archaea: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	65.6	230	11 Q9R149	G9r149 cavia porce
2	42	65.6	589	11 Q991L6	Q991L6 mus musculu
3	42	65.6	1450	13 Q9Y1B4	Q9Y1B4 cynops pyrr
4	42	65.6	1453	11 Q63079	Q63079 rattus norv
5	42	65.6	1461	4 Q76045	Q76045 homo sapien
6	40	62.5	113	11 Q8V172	Q8V172 cavia porce
7	40	62.5	347	6 Q9X125	Q9X125 bos taurus
8	40	62.5	475	10 Q80592	Q80592 arabidopsis
9	40	62.5	1160	4 Q14046	Q14046 homo sapien
10	40	62.5	1418	13 Q9W7R9	Q9W7R9 cynops pyrr
11	40	62.5	1418	13 Q63123	Q63123 cynops pyrr
12	40	62.5	1420	13 Q90W37	Q90W37 gallus gall
13	40	62.5	1442	11 Q62031	Q62031 mus musculu
14	40	62.5	1442	11 Q62033	Q62033 mus musculu
15	40	62.5	1459	11 Q62032	Q62032 mus musculu
16	40	62.5	1459	11 Q62032	Q62032 mus musculu

17	40	62.5	1487	4 Q14047	Q14047 homo sapien
18	40	62.5	1487	6 Q77753	Q77753 canis famil
19	39	60.9	211	2 Q9K17	Q9K17 ralsstoma s
20	38	59.4	373	2 Q9K17	Q9K17 yersinia en
21	38	59.4	373	2 Q9K17	Q9K17 yersinia en
22	38	59.4	695	10 Q82804	Q82804 arabidopsis
23	38	59.4	702	10 Q04419	Q04419 arabidopsis
24	38	59.4	809	13 Q93485	Q93485 oncorhynch
25	38	59.4	1449	13 Q910C0	Q910C0 oncorhynch
26	38	59.4	1647	16 Q9A180	Q9A180 streptococ
27	37.5	58.6	441	5 Q9U9P3	Q9U9P3 drosophila
28	37.5	58.6	953	5 Q9Y125	Q9Y125 drosophila
29	37	57.8	84	16 Q97550	Q97550 streptococ
30	37	57.8	328	4 Q9S061	Q9S061 homo sapien
31	37	57.8	463	5 Q8T0H4	Q8T0H4 drosophila
32	37	57.8	466	13 Q42434	Q42434 notophthalm
33	37	57.8	739	16 Q8XDP3	Q8XDP3 escherichia
34	37	57.8	771	16 Q8U724	Q8U724 agrobacteri
35	37	57.8	865	5 Q9V6K7	Q9V6K7 drosophila
36	37	57.8	1158	2 Q93EC2	Q93EC2 rhizobium l
37	37	57.8	1486	13 Q91717	Q91717 xenopus lae
38	36.5	57.0	365	5 Q85769	Q85769 caenorhabdi
39	36	56.2	100	4 Q9Y3P3	Q9Y3P3 homo sapien
40	36	56.2	149	16 Q915P4	Q915P4 pseudomonas
41	36	56.2	409	11 Q920V4	Q920V4 mus musculu
42	36	56.2	409	11 Q920V3	Q920V3 mus musculu
43	36	56.2	409	11 Q920V2	Q920V2 mus musculu
44	36	56.2	409	11 Q91VE9	Q91VE9 mus spicille
45	36	56.2	409	11 Q91V58	Q91V58 mus musculu
46	36	56.2	409	11 Q91V19	Q91V19 mus musculu
47	36	56.2	422	11 Q54905	Q54905 mus musculu
48	36	56.2	436	3 Q9P8N1	Q9P8N1 coriolus ve
49	36	56.2	441	4 Q96A84	Q96A84 homo sapien
50	36	56.2	491	2 Q9KWS5	Q9KWS5 pseudomonas
51	36	56.2	531	4 Q9BZG5	Q9BZG5 homo sapien
52	36	56.2	532	17 Q9YU22	Q9YU22 pyrococcus
53	36	56.2	539	4 Q9NUA2	Q9NUA2 homo sapien
54	36	56.2	542	4 Q9BZG6	Q9BZG6 homo sapien
55	36	56.2	544	4 Q9BZG7	Q9BZG7 homo sapien
56	36	56.2	555	4 Q9Y6M1	Q9Y6M1 homo sapien
57	36	56.2	598	16 Q8U6L9	Q8U6L9 agrobacteri
58	36	56.2	643	4 Q8TD91	Q8TD91 homo sapien
59	36	56.2	657	16 Q9ZL01	Q9ZL01 helicobacte
60	36	56.2	660	16 Q25414	Q25414 helicobacte
61	36	56.2	668	16 Q9ZKR5	Q9ZKR5 helicobacte
62	36	56.2	730	4 Q13771	Q13771 homo sapien
63	36	56.2	777	16 Q8ZLA2	Q8ZLA2 salmonella
64	36	56.2	777	16 Q8ZLA2	Q8ZLA2 salmonella
65	36	56.2	826	10 Q41059	Q41059 pisum sativ
66	36	56.2	830	10 Q04864	Q04864 sorghum tub
67	36	56.2	832	10 Q9M6P8	Q9M6P8 sorghum bic
68	36	56.2	847	10 Q9XISA	Q9XISA phaseolus v
69	36	56.2	852	10 Q08131	Q08131 manihot esc
70	36	56.2	906	4 Q9UN21	Q9UN21 homo sapien
71	36	56.2	1445	13 Q9J251	Q9J251 rana catesb
72	36	56.2	1447	13 Q91B91	Q91B91 xenopus lae
73	36	56.2	1779	5 Q9VMV4	Q9VMV4 drosophila
74	35	54.7	23	2 Q9KWN9	Q9KWN9 synecococc
75	35	54.7	89	2 Q54470	Q54470 staphylococ
76	35	54.7	182	5 Q9XWC2	Q9XWC2 caenorhabdi
77	35	54.7	259	11 Q99068	Q99068 mus musculu
78	35	54.7	369	5 Q21863	Q21863 caenorhabdi
79	35	54.7	393	11 Q9J9H7	Q9J9H7 mus musculu
80	35	54.7	396	11 Q9R084	Q9R084 rattus norv
81	35	54.7	396	11 Q9X0L3	Q9X0L3 rattus norv
82	35	54.7	422	4 Q9BZ09	Q9BZ09 homo sapien
83	35	54.7	422	4 Q43825	Q43825 homo sapien
84	35	54.7	437	16 Q97MC0	Q97MC0 clostridium
85	35	54.7	548	12 Q83880	Q83880 norway vir
86	35	54.7	548	12 Q68291	Q68291 human calli
87	35	54.7	562	12 Q90RA2	Q90RA2 tomato ring
88	35	54.7	562	12 Q90RA3	Q90RA3 tomato ring
89	35	54.7	564	12 Q91PA2	Q91PA2 tomato ring



236	33	51.6	972	16	P73273	P73273 synechocyst	309	32	50.0	533	16	Q8YND5	Q8YND5 anabaena sp
237	33	51.6	1011	3	Q9HEL5	Q9HEL5 neurospora	310	32	50.0	540	12	Q9IV49	Q9IV49 human calic
238	33	51.6	1225	3	Q9UJ98	Q9UJ98 homo sapien	311	32	50.0	540	12	Q9IBB0	Q9IBB0 norwalk-llk
239	33	51.6	1240	11	Q70576	Q70576 mus musculu	312	32	50.0	542	8	Q9BRU7	Q9BRU7 guillardi
240	33	51.6	1290	5	Q9NEA7	Q9NEA7 leishmania	313	32	50.0	546	3	P87306	P87306 schizosacch
241	33	51.6	1352	13	Q90YX0	Q90YX0 brachydanio	314	32	50.0	555	16	Q8VJ22	Q8VJ22 mycobacteri
242	33	51.6	1414	5	Q26634	Q26634 strongyloce	315	32	50.0	562	12	Q90RA5	Q90RA5 tomato ring
243	33	51.6	1675	12	Q90301	Q90301 rice tungro	316	32	50.0	572	5	Q20486	Q20486 caenorhabdl
244	33	51.6	1682	11	Q9QZK9	Q9QZK9 mus musculu	317	32	50.0	575	2	Q90RA5	Q90RA5 pseudomonas
245	33	51.6	1905	13	Q91823	Q91823 xenopus lae	318	32	50.0	576	2	Q9PD23	Q9PD23 pseudomonas
246	32.5	50.8	380	10	Q9S9V4	Q9S9V4 arabisopsis	319	32	50.0	592	6	Q9SRK3	Q9SRK3 canis famli
247	32.5	50.8	2471	3	Q59892	Q59892 candida alb	320	32	50.0	595	8	Q9M199	Q9M199 rhodospoete
248	32	50.0	77	9	Q9ZX19	Q9ZX19 bacterioph	321	32	50.0	595	8	Q9M199	Q9M199 typhloocte
249	32	50.0	82	10	Q8RZD2	Q8RZD2 oryza sativ	322	32	50.0	605	16	Q92E44	Q92E44 listeria in
250	32	50.0	103	2	Q9S336	Q9S336 prochloroco	323	32	50.0	620	2	Q9ANF0	Q9ANF0 bradyrhizob
251	32	50.0	111	17	Q8ZXK0	Q8ZXK0 pyrobaculum	324	32	50.0	622	4	Q96C38	Q96C38 homo sapien
252	32	50.0	118	2	Q9Z5E4	Q9Z5E4 phytoplasma	325	32	50.0	631	10	Q9LZU1	Q9LZU1 arabisopsis
253	32	50.0	127	2	Q9AKG6	Q9AKG6 nostoc punc	326	32	50.0	641	11	Q922K2	Q922K2 mus musculu
254	32	50.0	134	12	Q99BMS	Q99BMS human calic	327	32	50.0	771	5	Q19380	Q19380 caenorhabdl
255	32	50.0	175	17	Q9YD99	Q9YD99 aeropyrum p	328	32	50.0	803	2	Q93H85	Q93H85 bradyrhizob
256	32	50.0	186	13	Q9PSX5	Q9PSX5 paralicthy	329	32	50.0	806	2	Q45278	Q45278 typhloocte
257	32	50.0	187	11	Q9C0V9	Q9C0V9 mus musculu	330	32	50.0	812	5	Q06452	Q06452 ephydatia m
258	32	50.0	205	4	Q9H1F7	Q9H1F7 homo sapien	331	32	50.0	834	10	Q9RKE9	Q9RKE9 arabisopsis
259	32	50.0	211	10	Q9SLP2	Q9SLP2 carica papa	332	32	50.0	872	2	Q47114	Q47114 escherichia
260	32	50.0	214	3	Q9HEZ5	Q9HEZ5 neurospora	333	32	50.0	873	4	Q9UMF9	Q9UMF9 homo sapien
261	32	50.0	220	16	Q9KQ08	Q9KQ08 vibrio chol	334	32	50.0	873	4	Q96M72	Q96M72 homo sapien
262	32	50.0	221	16	Q97H92	Q97H92 clostridium	335	32	50.0	896	4	Q96SFO	Q96SFO homo sapien
263	32	50.0	221	16	Q9RJCS	Q9RJCS streptomyce	336	32	50.0	901	13	Q921B1	Q921B1 mus musculu
264	32	50.0	243	5	Q9BKN6	Q9BKN6 strongyloce	337	32	50.0	901	13	Q91948	Q91948 xenopus sp.
265	32	50.0	247	5	Q9BHM4	Q9BHM4 strongyloce	338	32	50.0	901	16	Q9PH65	Q9PH65 xylella fas
266	32	50.0	247	5	Q9BKN5	Q9BKN5 strongyloce	339	32	50.0	902	10	Q9M364	Q9M364 arabisopsis
267	32	50.0	249	16	Q9BE77	Q9BE77 rhizobium 1	340	32	50.0	904	13	Q91977	Q91977 xenopus lae
268	32	50.0	250	5	Q9BKN4	Q9BKN4 strongyloce	341	32	50.0	1040	5	Q925I8	Q925I8 mus musculu
269	32	50.0	252	7	Q31369	Q31369 brachydanio	342	32	50.0	1060	5	Q9VSR1	Q9VSR1 drosophila
270	32	50.0	268	5	Q17030	Q17030 anopheles g	343	32	50.0	1064	10	Q041S5	Q041S5 arabisopsis
271	32	50.0	276	16	Q8XP0	Q8XP0 talstonia s	344	32	50.0	1064	10	Q041S6	Q041S6 arabisopsis
272	32	50.0	277	10	Q94L68	Q94L68 oryza sativ	345	32	50.0	1064	10	Q04884	Q04884 arabisopsis
273	32	50.0	293	17	Q9YDC3	Q9YDC3 aeropyrum p	346	32	50.0	1064	10	Q9SMW4	Q9SMW4 arabisopsis
274	32	50.0	297	16	Q8XCY2	Q8XCY2 escherichia	347	32	50.0	1093	4	Q9P2E7	Q9P2E7 homo sapien
275	32	50.0	304	16	Q91798	Q91798 pseudomonas	348	32	50.0	1129	5	Q93342	Q93342 caenorhabdl
276	32	50.0	306	16	Q8U5H3	Q8U5H3 agrobacteri	349	32	50.0	1140	11	Q61434	Q61434 mus musculu
277	32	50.0	307	5	Q18327	Q18327 caenorhabdl	350	32	50.0	1140	16	Q9CNS4	Q9CNS4 pasteurella
278	32	50.0	308	16	Q9X7E8	Q9X7E8 mycobacteri	351	32	50.0	1256	11	Q99M76	Q99M76 rattus norv
279	32	50.0	309	8	Q03140	Q03140 acorus gram	352	32	50.0	1264	12	Q89230	Q89230 variola vir
280	32	50.0	309	16	Q92U62	Q92U62 rhizobium m	353	32	50.0	1336	10	Q810B3	Q810B3 arabisopsis
281	32	50.0	310	16	Q9S2J4	Q9S2J4 streptomyce	354	32	50.0	1347	4	Q960B3	Q960B3 homo sapien
282	32	50.0	311	2	Q30422	Q30422 caldocellum	355	32	50.0	1378	5	Q97405	Q97405 halloia di
283	32	50.0	321	12	Q9U454	Q9U454 human rotav	356	32	50.0	1491	13	Q91718	Q91718 xenopus lae
284	32	50.0	330	10	Q946D5	Q946D5 arabisopsis	357	32	50.0	1497	11	Q61431	Q61431 mus musculu
285	32	50.0	333	16	Q9HMI3	Q9HMI3 pseudomonas	358	32	50.0	1524	10	Q8WSF5	Q8WSF5 oryza sativ
286	32	50.0	345	16	Q9CNK9	Q9CNK9 pasteurella	359	32	50.0	1551	5	Q9VMP3	Q9VMP3 drosophila
287	32	50.0	361	3	Q9UT98	Q9UT98 schizosacch	360	32	50.0	1573	5	Q95V18	Q95V18 drosophila
288	32	50.0	370	3	P78821	P78821 schizosacch	361	32	50.0	1648	11	Q9ESB7	Q9ESB7 mus musculu
289	32	50.0	376	11	Q9JK15	Q9JK15 rattus norv	362	32	50.0	1648	11	Q9Q205	Q9Q205 rice tungro
290	32	50.0	380	5	Q07138	Q07138 microclonia	363	32	50.0	1674	12	Q91DT3	Q91DT3 white spot
291	32	50.0	381	10	Q9LDD5	Q9LDD5 arabisopsis	364	32	50.0	1684	12	Q91LK9	Q91LK9 white spot
292	32	50.0	382	5	Q9N3T2	Q9N3T2 caenorhabdl	365	32	50.0	1701	3	Q8WZK8	Q8WZK8 white spot
293	32	50.0	390	10	Q9LJZ4	Q9LJZ4 oryza sativ	366	32	50.0	1774	11	Q62001	Q62001 mus musculu
294	32	50.0	399	16	Q98IP0	Q98IP0 rhizobium 1	367	32	50.0	1869	12	Q8V2H2	Q8V2H2 camelpox vi
295	32	50.0	403	4	Q8BEL7	Q8BEL7 homo sapien	368	32	50.0	1869	12	Q8QPZ7	Q8QPZ7 camelpox vi
296	32	50.0	403	16	Q99WH0	Q99WH0 staphylococ	369	32	50.0	1896	12	Q89117	Q89117 variola vir
297	32	50.0	421	16	Q84091	Q84091 chlamydia t	370	32	50.0	1896	12	Q89192	Q89192 variola vir
298	32	50.0	425	2	Q9RMB0	Q9RMB0 bacillus ce	371	32	50.0	1896	12	Q89096	Q89096 variola vir
299	32	50.0	443	4	Q9HGL3	Q9HGL3 schizosacch	372	32	50.0	1897	12	Q89406	Q89406 variola maj
300	32	50.0	446	4	Q9Y2B1	Q9Y2B1 homo sapien	373	32	50.0	1919	12	Q80MM9	Q80MM9 cowpox viru
301	32	50.0	448	10	Q9MAT6	Q9MAT6 arabisopsis	374	32	50.0	1924	12	Q9JFS2	Q9JFS2 ecclimella
302	32	50.0	461	10	Q9FTD4	Q9FTD4 oryza sativ	375	32	50.0	1933	12	Q72759	Q72759 cowpox viru
303	32	50.0	467	16	Q8YWE0	Q8YWE0 anabaena sp	376	32	50.0	1958	11	P70276	P70276 mus musculu
304	32	50.0	468	16	Q9L0P9	Q9L0P9 streptomyce	377	32	50.0	2090	3	Q9P6X4	Q9P6X4 neurospora
305	32	50.0	478	17	Q28457	Q28457 archaeoglob	378	32	50.0	2684	10	Q9LNF9	Q9LNF9 arabisopsis
306	32	50.0	487	16	Q9A7V2	Q9A7V2 caulobacter	379	32	50.0	4099	10	Q9C7Z6	Q9C7Z6 arabisopsis
307	32	50.0	492	15	Q88987	Q88987 visna virus	380	32	50.0	6239	2	Q9S0R7	Q9S0R7 streptomyce
308	32	50.0	527	16	Q97H21	Q97H21 clostridium	381	32	50.0				

382	31.5	49.2	580	16	Q8XRM3	Q8XRM3 talstonia s
383	31.5	49.2	632	17	Q9V229	Q9V229 pyrococcus
384	31.5	49.2	641	17	Q8TZN5	Q8TZN5 pyrococcus
385	31.5	49.2	788	10	Q94ZH4	Q94ZH4 oryza sativ
386	31.5	49.2	819	11	Q923X3	Q923X3 cricetus
387	31.5	49.2	1900	12	Q9WTF1	Q9WTF1 sindbis-11k
388	31.5	49.2	1901	12	Q884Z9	Q884Z9 sindbis-11k
389	31.5	49.2	2500	12	Q884J3	Q884J3 sindbis-11k
390	31.5	49.2	2513	12	Q87644	Q87644 sindbis vir
391	31.5	49.2	2517	12	Q9YXK6	Q9YXK6 sindbis-11k
392	31	48.4	51	16	Q8VWC2	Q8VWC2 mycobacteri
393	31	48.4	79	9	Q9XJF0	Q9XJF0 lactococcus
394	31	48.4	79	9	Q94M89	Q94M89 lactococcus
395	31	48.4	81	2	Q9E0Z3	Q9E0Z3 escherichia
396	31	48.4	98	8	Q995Z5	Q995Z5 cyprinella
397	31	48.4	99	11	Q91WC4	Q91WC4 mus musculu
398	31	48.4	108	16	Q8Z6L4	Q8Z6L4 salmonella
399	31	48.4	108	16	Q84944	Q84944 salmonella
400	31	48.4	116	4	Q8TDC8	Q8TDC8 homo sapien
401	31	48.4	134	16	Q92KX5	Q92KX5 rhizobium m
402	31	48.4	136	11	Q923N9	Q923N9 mus musculu
403	31	48.4	139	9	Q9T1L4	Q9T1L4 bacterioph
404	31	48.4	139	16	Q8X697	Q8X697 escherichia
405	31	48.4	146	16	Q92PML	Q92PML rhizobium m
406	31	48.4	150	6	Q77789	Q77789 bos taurus
407	31	48.4	155	11	Q9D9L7	Q9D9L7 mus musculu
408	31	48.4	157	2	Q9X4L1	Q9X4L1 paracoccus
409	31	48.4	157	2	P95314	P95314 mycobacteri
410	31	48.4	159	4	Q8WMD6	Q8WMD6 homo sapien
411	31	48.4	161	10	Q8WVU9	Q8WVU9 narcissus p
412	31	48.4	163	16	Q9HVA1	Q9HVA1 pseudomonas
413	31	48.4	173	4	Q96CZ6	Q96CZ6 homo sapien
414	31	48.4	184	4	Q96T43	Q96T43 homo sapien
415	31	48.4	184	17	Q9YEH5	Q9YEH5 aeropyrum p
416	31	48.4	188	16	Q97PZ8	Q97PZ8 streptococ
417	31	48.4	192	16	Q8VJUR4	Q8VJUR4 mycobacteri
418	31	48.4	197	16	Q9HT85	Q9HT85 pseudomonas
419	31	48.4	198	2	Q9KML1	Q9KML1 staphylococ
420	31	48.4	206	2	Q925Y8	Q925Y8 corynebacte
421	31	48.4	208	4	Q9U777	Q9U777 homo sapien
422	31	48.4	208	4	Q961Y6	Q961Y6 homo sapien
423	31	48.4	209	10	Q23485	Q23485 arabidopsi
424	31	48.4	213	4	Q9UBM0	Q9UBM0 homo sapien
425	31	48.4	218	16	Q9KXY6	Q9KXY6 streptomyc
426	31	48.4	219	5	Q96537	Q96537 allocentrot
427	31	48.4	219	5	Q9TYD4	Q9TYD4 allocentrot
428	31	48.4	220	4	Q9NS60	Q9NS60 homo sapien
429	31	48.4	220	4	Q9HD11	Q9HD11 homo sapien
430	31	48.4	220	6	Q8WNO4	Q8WNO4 mustela put
431	31	48.4	220	11	Q9JUM6	Q9JUM6 rattus norv
432	31	48.4	220	11	Q9JUF8	Q9JUF8 mus musculu
433	31	48.4	220	11	Q9J121	Q9J121 rattus norv
434	31	48.4	220	11	Q8VHNA	Q8VHNA mus musculu
435	31	48.4	221	5	Q95S19	Q95S19 drosophila
436	31	48.4	225	4	Q9HON4	Q9HON4 homo sapien
437	31	48.4	225	4	Q96K86	Q96K86 homo sapien
438	31	48.4	227	4	Q9HD10	Q9HD10 homo sapien
439	31	48.4	228	2	Q9WZ44	Q9WZ44 pseudomonas
440	31	48.4	228	5	Q96534	Q96534 strongyloce
441	31	48.4	232	5	Q25363	Q25363 loligo opal
442	31	48.4	232	5	Q25367	Q25367 loligo opal
443	31	48.4	235	5	Q96535	Q96535 strongyloce
444	31	48.4	236	5	Q96538	Q96538 hemientrot
445	31	48.4	236	5	Q97Y05	Q97Y05 hemientrot
446	31	48.4	237	16	Q9AHJ5	Q9AHJ5 escherichia
447	31	48.4	242	2	Q9E290	Q9E290 lactococcus
448	31	48.4	243	10	Q8RXT9	Q8RXT9 arabidopsi
449	31	48.4	245	2	Q93TM9	Q93TM9 nostoc punc
450	31	48.4	252	4	Q9NZ11	Q9NZ11 homo sapien
451	31	48.4	252	5	Q9BKM8	Q9BKM8 strongyloce
452	31	48.4	252	6	Q8WN05	Q8WN05 mustela put
453	31	48.4	252	11	Q9J122	Q9J122 rattus norv
454	31	48.4	252	11	Q8VHNS	Q8VHNS mus musculu

455	31	48.4	255	5	Q9BKM9	Q9BKM9 strongyloce
456	31	48.4	256	5	Q9BKN1	Q9BKN1 strongyloce
457	31	48.4	256	11	Q923P1	Q923P1 mus musculu
458	31	48.4	259	10	Q9SEW5	Q9SEW5 vicia faba
459	31	48.4	260	16	Q06345	Q06345 mycobacteri
460	31	48.4	262	5	Q9BRN0	Q9BRN0 strongyloce
461	31	48.4	263	10	Q9SEW7	Q9SEW7 vicia faba
462	31	48.4	266	5	P91608	P91608 drosophila
463	31	48.4	267	10	Q9SCR9	Q9SCR9 arabidopsi
464	31	48.4	269	16	Q8X613	Q8X613 escherichia
465	31	48.4	270	4	Q9NS61	Q9NS61 homo sapien
466	31	48.4	270	4	Q96T41	Q96T41 homo sapien
467	31	48.4	270	5	Q9BKN2	Q9BKN2 strongyloce
468	31	48.4	270	5	Q9T206	Q9T206 strongyloce
469	31	48.4	270	5	Q96536	Q96536 strongyloce
470	31	48.4	270	6	Q8WNO3	Q8WNO3 mustela put
471	31	48.4	270	11	Q9JMS9	Q9JMS9 rattus norv
472	31	48.4	270	11	Q9JUF6	Q9JUF6 mus musculu
473	31	48.4	270	11	Q9J123	Q9J123 rattus norv
474	31	48.4	272	11	Q8VHNE	Q8VHNE mus musculu
475	31	48.4	272	10	Q943Z4	Q943Z4 arabidopsi
476	31	48.4	277	5	Q9BKN3	Q9BKN3 strongyloce
477	31	48.4	283	16	Q9RX60	Q9RX60 delnecoccus
478	31	48.4	285	4	Q96T44	Q96T44 homo sapien
479	31	48.4	289	16	Q8Y1T5	Q8Y1T5 talstonia s
480	31	48.4	290	2	Q8RLR1	Q8RLR1 azocarc to
481	31	48.4	309	2	Q9F694	Q9F694 streptococ
482	31	48.4	311	16	Q9A3C4	Q9A3C4 caulobacter
483	31	48.4	312	2	Q9F693	Q9F693 streptococ
484	31	48.4	313	5	Q77284	Q77284 drosophila
485	31	48.4	316	16	Q8XU25	Q8XU25 arabidopsi
486	31	48.4	317	16	Q9KWL1	Q9KWL1 vibrio chol
487	31	48.4	317	16	Q8YPM2	Q8YPM2 arabidopsi
488	31	48.4	319	12	Q88700	Q88700 hepatitis c
489	31	48.4	319	17	Q8ZTRA3	Q8ZTRA3 pyrobaculum
490	31	48.4	330	16	Q9EMF0	Q9EMF0 streptomyc
491	31	48.4	333	17	Q8TL89	Q8TL89 methanosarc
492	31	48.4	334	15	Q9TUB9	Q9TUB9 human immun
493	31	48.4	339	10	Q9SK61	Q9SK61 arabidopsi
494	31	48.4	342	2	Q8RT50	Q8RT50 pseudomonas
495	31	48.4	342	16	Q9AAV9	Q9AAV9 pseudomonas
496	31	48.4	343	16	Q92ZV7	Q92ZV7 rhizobium m
497	31	48.4	345	5	Q95YPT7	Q95YPT7 leishmania
498	31	48.4	346	5	Q96FM2	Q96FM2 ceenorhabdi
499	31	48.4	348	16	Q9KHM4	Q9KHM4 streptococ
500	31	48.4	349	10	Q9SM53	Q9SM53 arabidopsi

## ALIGNMENTS

## RESULT 1

Q9R149 PRELIMINARY; PRT; 230 AA.

AC Q9R149;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 21, Last annotation update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Pro-alpha-1 type 1 collagen (Fragment).

OS Cavia porcellus (Guinea pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Hystriicognathi; Cavidae; Cavia.

OX NCBI\_TaxID=10141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HARLEY; TISSUE=TRACHEA;

RA Morishima Y., Uchida Y., Nomura A., Ishii Y., Sakamoto T.,

RA Sekizawa K.;

RT "Guinea pig" pro-alpha-1 type 1 collagen expression in injured tracheal

RT epithelium.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF169346; AAD49346.1; -

InterPro; IPR000087; Collagen.

DR Pfam: PF01391; Collagen; 4.  
 KW Collagen.  
 FT NON\_TER 1  
 FT NON\_TER 230  
 SO SEQUENCE 230 AA; 20425 MW; 1A465F92779D9A71 CRC64;

Query Match 65.6%; Score 42; DB 11; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 4.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 82 POGIAGOR 89

## RESULT 2

O99LL6 PRELIMINARY; PRT; 589 AA.

AC O99LL6; 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE Hypothetical 58.8 kDa protein (Fragment).

GN COL1A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC003198; AAH03198.1;  
 DR MGI: 88467; COL1A1.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 6.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR KEGG: Collagen; Hypothetical protein.  
 FT NON\_TER 1  
 SO SEQUENCE 589 AA; 58805 MW; 81847495E505CEF CRC64;

Query Match 65.6%; Score 42; DB 11; Length 589;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 76 POGIAGOR 83

## RESULT 3

O9YIB4 PRELIMINARY; PRT; 1450 AA.

AC O9YIB4; 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DE Alpha 1 type I collagen.

DE Cynops pyrrhogaster (Japanese common newt).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandridae; Cynops.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-REGENERATE FORELIMBS;  
 RX MEDLINE-99407244; PubMed-10474166;  
 RA Asahina K., Obara M., Yoshizato K.;  
 RT "Expression of genes of type I and type II collagen in the formation  
 and development of the blastema of regenerating newt limb.";  
 Dev. Dyn. 216:59-71(1999).

DR EMBL: AB015438; BAA36973.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Collagen; 1.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWF; 1.  
 DR PROSITE: PS01208; VWF; UNKNOWN\_1.  
 DR COLLAGEN.  
 SO SEQUENCE 1450 AA; 137564 MW; ABF8A74841B8767C CRC64;

Query Match 65.6%; Score 42; DB 13; Length 1450;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 937 POGIAGOR 944

## RESULT 4

O63079 PRELIMINARY; PRT; 1453 AA.

AC O63079; 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DE Collagen alpha1 (Fragment).  
 DE Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE OF 1-1092 FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TOOTH;  
 RA Brandsten C., Lundmark C., Christenson C., Hammarstrom L., Wurtz T.;  
 RT "Expression of Collagen alpha1(I) mRNA variants during tooth and bone  
 formation in the Rat.";  
 J. Dent. Res. 0:0-0(0).  
 FT NON\_TER 1  
 SO SEQUENCE 1453 AA; 137887 MW; E6896BDC19A4A1D8 CRC64;

Query Match 65.6%; Score 42; DB 11; Length 1453;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 940 POGIAGOR 947

## RESULT 5

O76045

ID 076045 PRELIMINARY; PRT; 1461 AA.  
AC 076045;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Pro alpha 1(I) collagen.  
GN COL1A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN SEQUENCE FROM N.A.  
RP MEDLINE=85130970; PubMed=2857713;  
RX Chu M.L., de Wet W., Bernard M., Ramirez F.;  
RT "Fine structural analysis of the human pro-alpha 1(I) collagen gene.  
RT Promoter structure, AluI repeats, and polymorphic transcripts.";  
RL J. Biol. Chem. 260:2315-2320(1985).  
RN SEQUENCE FROM N.A.  
RP MEDLINE=88329734; PubMed=2843432;  
RX D'Allesio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;  
RT "Complete nucleotide sequence of the region encompassing the first  
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";  
RL Gene 67:105-115(1988).  
RN SEQUENCE FROM N.A.  
RP MEDLINE=9138770; PubMed=1995349;  
RX Maatta A., Bornstein P., Penttinen R.P.;  
RT "Highly conserved sequences in the 3'-untranslated region of the  
RT COL1A1 gene bind cell-specific nuclear proteins.";  
RL FEBS Lett. 279:9-13(1991).  
RN SEQUENCE FROM N.A.  
RP MEDLINE=92157916; PubMed=1787829;  
RX Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,  
RA Olsen A.S., Prockop D.J.;  
RT "Completion of the last half of the structure of the human gene for  
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";  
RL Matrix 11:375-379(1991).  
RN SEQUENCE FROM N.A.  
RP MEDLINE=96107942; PubMed=9443882;  
RX Korhko J., Ala-Korhko L., De Paape A., Nuytlinck L., Earley J.,  
RA Prockop D.J.;  
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and  
RT scanning by conformation-sensitive gel electrophoresis identifies only  
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:  
RT identification of common sequences of null-allele mutations.";  
RL Am. J. Hum. Genet. 62:98-110(1998).  
RN SEQUENCE FROM N.A.  
RA Ala-Korhko L., Earley J.J., Nuytlinck L., DePaape A., Prockop D.J.,  
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF017178; AAB94054.2; -  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR000885; Fib\_collagen\_C.  
DR InterPro: IPR001007; WVF\_C.  
DR Pfam: PF01410; COLFI; 1.  
DR Pfam: PF01391; Collagen; 18.  
DR Pfam: PF00093; WVC; 1.  
DR ProDom: PD000007; Collagen; 2.  
DR ProDom: PD002078; Fib\_collagen\_C; 1.  
DR SMART: SM00038; COLFI; 1.

DR SMART: SM00214; WVC; 1.  
DR PROSITE: PS01208; WVFC; 1.  
KW Collagen.  
SQ SEQUENCE 1461 AA; 138630 MW; 9ACFDE30EA78E21 CRC64;  
Query Match 65.6%; Score 42; DB 4; Length 1461;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 POGIACOR 9  
DB 948 POGIACOR 955  
RESULT 6  
ID 08V172 PRELIMINARY; PRT; 113 AA.  
AC 08V172;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Collagen type II (Fragment).  
OC Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.  
OX NCBI\_TaxID=10141;  
RN SEQUENCE FROM N.A.  
RP STRAIN=HARTLEY;  
RA Huebner J.L., Clark A.E., Kraus V.B., Otterness I.G.;  
RT "Collagen type II in the guinea pig.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF152862; AAL55558.1; -  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen; 1.  
DR ProDom: PD000007; Collagen; 1.  
FT NON\_TER 1 113  
FT NON\_TER 1 113  
SQ SEQUENCE 113 AA; 10284 MW; F7861901127A9BCE CRC64;  
Query Match 62.5%; Score 40; DB 11; Length 113;  
Best Local Similarity 87.5%; Pred. No. 5.4;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 POGIACOR 9  
DB 93 POGIACOR 100  
RESULT 7  
ID 09XT25 PRELIMINARY; PRT; 347 AA.  
AC 09XT25;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Type II collagen cyanogen bromide fragment CB10 (Fragment).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN SEQUENCE FROM N.A.  
RP TISSUE=CARTILAGE;  
RA Tang B., Chiang T.M., Brand D.D., Gumanovskaya M.L., Stuart J.M.,  
RA Kang A.H., Myers L.K.;  
RT "Molecular Definition and Characterization of Recombinant Bovine CB8  
RT And CB10: Immunogenicity and Arthritisogenicity.";  
RL J. Clin. Immunol. 0:0-0(1999).  
DR EMBL: AF138883; AAD42346.1; -  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen; 6.

DR ProDom: PD000007; Collagen; 2.  
 KW Collagen.  
 FT NON\_TER 1 1  
 SO SEQUENCE 347 AA; 31085 MW; 5D1C0AF34089DF6 CRC64;

Query Match  
 Best Local Similarity 62.5%; Score 40; DB 6; Length 347;  
 Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 223 POGIAGOR 230

RESULT 8  
 ID 080592 PRELIMINARY; PRT: 475 AA.  
 AC 080592;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative amino acid perase.  
 GN T2711.3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Federle N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
 RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,  
 RA O.J.O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,  
 RA Yu G., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC004122; AAC34329.1;  
 DR InterPro: IPR002422; AA/rel\_primease2.  
 DR Pfam: PF01490; AA\_trans; 1.  
 SO SEQUENCE 475 AA; 51815 MW; B6DE571E2FE7F92 CRC64;

Query Match  
 Best Local Similarity 58.3%; Score 40; DB 10; Length 475;  
 Pred. No. 25;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 SPOGIAGORFN 12  
 DB 87 SPDSITGRN 98

RESULT 9  
 ID 014046 PRELIMINARY; PRT: 1160 AA.  
 AC 014046;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE COL2A1 protein precursor (Fragment).  
 GN COL2A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CARTILAGE;  
 RA MEDLINE-90026318; PubMed-2803268;  
 RA Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;  
 RT Structure of cDNA clones coding for human type II procollagen. The  
 RT alpha 1(I) chain is more similar to the alpha 1(I) chain than two  
 RT other alpha chains of fibrillar collagens.";  
 RL Biochem. J. 262:521-528(1989).

DR EMBL: X16711; CAA34683.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Collagen; 4.  
 KW Collagen; signal.  
 FT SIGNAL 1 23  
 FT CHAIN 113 >1160  
 FT NON\_TER 1160 1160  
 SO SEQUENCE 1160 AA; 105630 MW; A7F0523B856C8639 CRC64;

Query Match  
 Best Local Similarity 62.5%; Score 40; DB 4; Length 1160;  
 Pred. No. 63;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 904 POGIAGOR 911

RESULT 10  
 ID 028396 PRELIMINARY; PRT: 1418 AA.  
 AC 028396;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Type II collagen.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_Taxid=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Richardson D.W., Dodge G.R.;  
 RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RA MacLeod J.N., Fubini S.L., Gu D.N., Tetreault J.W., Todhunter R.J.;  
 RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U62528; AAB05773.1;  
 DR EMBL; AF040638; AAB96768.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Collagen; 4.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 KW Collagen.  
 SO SEQUENCE 1418 AA; 134343 MW; 115FCD19EB8696A3 CRC64;

Query Match  
 Best Local Similarity 62.5%; Score 40; DB 6; Length 1418;  
 Pred. No. 78;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 904 POGIAGOR 911

RESULT 11  
 ID 09W789 PRELIMINARY; PRT: 1418 AA.  
 AC 09W789;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Alpha type II collagen.  
 OS Cynops pyrrhogaster (Japanese common newt).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae; Cynops.  
 OX NCBI\_Taxid=8330;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=99407244; PubMed=10474166;  
 RA Asahina K., Obara M., Yoshizato K.;  
 RT "Expression of genes of type I and type II collagen in the formation  
 and development of the blastema of regenerating newt limb.";  
 RL Dev. Dyn. 216:59-71(1999).  
 DR EMBL: AB022046; BA82043.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Collagen; 7.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 KW Collagen.  
 SQ SEQUENCE 1418 AA; 135067 MW; C19A6E601A2A717E CRC64;

Query Match 62.5%; Score 40; DB 13; Length 1418;  
 Best Local Similarity 87.5%; Pred. No. 78;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 POGIAGOR 9  
 DB 904 POGIAGOR 911

## RESULT 12

063123 PRELIMINARY; PRT; 1419 AA.  
 AC 063123; 063565;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Collagen alpha 1 type II (T1 mRNA).  
 GN T1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BONE FRACTURE CALLUS;  
 RA Urahe K., Sarkar G., Bolander M.E.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1372-1419 FROM N.A.  
 RA Wurtz T., Brandsten C., Lundmark C., Christersson C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 370-422 FROM N.A.  
 RC STRAIN-DA; TISSUE-CARTILAGE;  
 RX MEDLINE=94321934; PubMed=8046350;  
 RA Michaelson E., Malmstrom V., Reis S., Engstrom A., Burkhardt H.,  
 RA Holmdahl R.;  
 RT "T cell recognition of carbohydrates on type II collagen.";  
 RL J. Exp. Med. 180:745-749(1994).  
 DR EMBL: I48440; AAY9780.1;  
 DR EMBL: AJ224879; CA12179.1;  
 DR EMBL: X79816; CA56213.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Collagen; 4.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 KW Collagen.  
 SQ SEQUENCE 1419 AA; 134570 MW; B7C63B77819CE50B CRC64;

Query Match 62.5%; Score 40; DB 11; Length 1419;  
 Best Local Similarity 87.5%; Pred. No. 78;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 POGIAGOR 9  
 DB 905 POGIAGOR 912

## RESULT 13

090W37 PRELIMINARY; PRT; 1420 AA.  
 AC 090W37;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Alpha 1 type IIA collagen precursor.  
 GN COL2A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OX Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=STERNAL;  
 RA Caixia X., Yongzhi X., Siqi G., Yiyang S.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY046949; AAK98621.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR ProDom: PD000007; Collagen; 6.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1 25  
 SQ SEQUENCE 1420 AA; 134999 MW; 88D9AAB17F214FF5 CRC64;

Query Match 62.5%; Score 40; DB 13; Length 1420;  
 Best Local Similarity 87.5%; Pred. No. 78;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 POGIAGOR 9  
 DB 906 POGIAGOR 913

## RESULT 14

062031 PRELIMINARY; PRT; 1442 AA.  
 AC 062031;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Pro-alpha-1 type II collagen.  
 GN COL2A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57/BLACK;  
 RX MEDLINE=91358489; PubMed=1885613;  
 RA Metzaranta M., Roman D., de Crombrughe B., Vuorio E.;  
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon  
 structure, and alternative splicing.";  
 RL J. Biol. Chem. 266:16862-16869(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57/BLACK;  
 RA Vuorio E.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: M65161; AAG68099.1;  
 DR MGD: MGI:88452; Col2a1.  
 DR InterPro: IPR000087; Collagen.



DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 17.  
 DR Pfam: PF00093; VWC; 1.  
 DR ProDom: PD000007; Collagen; 5.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWC; 1.  
 DR PROSITE: PS01208; VWF; 1.  
 DR Collagen.  
 SO SEQUENCE 1442 AA; 137911 MW; D3A3274493C8821C CRC64;

Query Match 62.5%; Score 40; DB 11; Length 1442;  
 Best Local Similarity 87.5%; Pred. No. 80;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 928 POGIAGOR 935

## RESULT 15

062033 PRELIMINARY; PRT; 1442 AA.  
 ID 062033;  
 AC 062033;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Pro-alpha-1 type II collagen.  
 OS COL2A1  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57/BLACK;  
 RX MEDLINE=91358489; PubMed=1885613;  
 RA Metzaranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon  
 structure, and alternative splicing."  
 RT J. Biol. Chem. 266:16862-16869(1991).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=C57/BLACK;  
 RA Vuorio E.;  
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: M5161; AAA68101.1; -  
 DR MGD: MGI:88452; Col2a1.  
 DR InterPro: IPR000887; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 17.  
 DR Pfam: PF00093; VWC; 1.  
 DR ProDom: PD000007; Collagen; 5.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWC; 1.  
 DR PROSITE: PS01208; VWF; 1.  
 DR Collagen.  
 SO SEQUENCE 1442 AA; 137829 MW; F0E77C11BCAFA93B CRC64;

Query Match 62.5%; Score 40; DB 11; Length 1442;  
 Best Local Similarity 87.5%; Pred. No. 80;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 928 POGIAGOR 935

## RESULT 16

062032 PRELIMINARY; PRT; 1459 AA.  
 ID 062032;  
 AC 062032;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Pro-alpha-1 type II collagen.  
 OS COL2A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57/BLACK;  
 RX MEDLINE=91358489; PubMed=1885613;  
 RA Metzaranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon  
 structure, and alternative splicing."  
 RT J. Biol. Chem. 266:16862-16869(1991).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=C57/BLACK;  
 RA Vuorio E.;  
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: M5161; AAA68101.1; -  
 DR MGD: MGI:88452; Col2a1.  
 DR InterPro: IPR000887; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR Pfam: PF00093; VWC; 1.  
 DR ProDom: PD000007; Collagen; 5.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWC; 1.  
 DR PROSITE: PS01208; VWF; 1.  
 DR Collagen.  
 SO SEQUENCE 1459 AA; 139071 MW; A09D24BF7357C827 CRC64;

Query Match 62.5%; Score 40; DB 11; Length 1459;  
 Best Local Similarity 87.5%; Pred. No. 81;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 DB 945 POGIAGOR 952

## RESULT 17

014047 PRELIMINARY; PRT; 1487 AA.  
 ID 014047;  
 AC 014047;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Alpha-1 type II collagen.  
 OS COL2A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BLOOD;  
 RX MEDLINE=85190534; PubMed=3857598;  
 RA Cheah K.S., Stoker N.G., Griffin J.R., Grosveid F.G., Solomon E.;  
 RT "Identification and characterization of the human type II collagen  
 gene (COL2A1)."  
 RT Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).  
 RL [2]

RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE=90026318; PubMed=2803268;  
RA Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;  
RT "Structure of cDNA clones coding for human type II procollagen. The  
RT alpha 1(II) chain is more similar to the alpha 1(I) chain than two  
RT other alpha chains of fibrillar collagens.";  
RL Biochem. J. 262:521-528(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE=89325561; PubMed=2753125;  
RA Vlkula M., Peltonen L.;  
RT "Structural analyses of the polymorphic area in type II collagen  
RT gene.";  
RL FEBS Lett. 250:171-174(1989).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE=91184811; PubMed=2081599;  
RA Ryan M.C., Sieraski M., Sandell L.J.;  
RT "The human type II procollagen gene: identification of an additional  
RT protein-coding domain and location of potential regulatory sequences  
RT in the promoter and first intron.";  
RL Genomics 8:41-48(1990).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE=91153296; PubMed=1999183;  
RA Huang M.C., Seyer J.M., Thompson J.P., Spinella D.G., Cheah K.S.,  
RT "Genomic organization of the human procollagen alpha 1(II) collagen  
RT gene.";  
RL Eur. J. Biochem. 195:593-600(1991).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE=92344585;  
RA Vlkula M., Metsaranta M., Syvanen A.C., Ala-Kokko I., Vuorio E.,  
RT "Structural analysis of the regulatory elements of the type-II  
RT procollagen gene. Conservation of promoter and first intron sequences  
RT between human and mouse.";  
RL Biochem. J. 285:0-0(0).  
RN [7]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE=97104294; PubMed=8948452;  
RA Ala-Kokko I., Kvist A.P., Metsaranta M., Kivirikko K.I.,  
RT "Conservation of the sizes of 5' introns and over 100 intronic  
RT sequences for the binding of common transcription factors in the human  
RT and mouse genes for type II procollagen (COL2A1).";  
RL Biochem. J. 308:0-0(0).  
RN EMBL: L10347; AAC41772.1; -;  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR000885; Fib\_collagen\_C.  
DR InterPro: IPR001007; VWF\_C.  
DR Pfam: PF01410; COLFI; 1.  
DR Pfam: PF01391; Collagen; 18.  
DR Pfam: PF00093; VWC; 1.  
DR ProDom: PD000007; Collagen; 4.  
DR ProDom: PD002078; Fib\_collagen\_C; 1.  
DR SMART: SM00038; COLFI; 1.  
DR SMART: SM00214; VWC; 1.  
DR PROSITE: PS01208; VWF; 1.  
KM Collagen.  
SQ SEQUENCE 1487 AA; 141772 MW; 0B7E79D46BDAFA97 CRC64;

QY 2 POGIAGOR 9  
Db 973 POGIAGOR 980  
RESULT 18  
ID 077753 PRELIMINARY; PRT; 1487 AA.  
AC 077753;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Type IIA procollagen.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96340920; PubMed=9676231;  
RA Du F., Acland G.M., Ray J.;  
RT "Differential splicing of type II procollagen mRNA in canine retina.";  
RL Anim. Biotechnol. 9:15-20(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20480698; PubMed=11024291;  
RA Du F., Acland G.M., Ray J.;  
RT "Cloning and expression of type II collagen mRNA: evaluation as a  
RT candidate for canine ocular skeletal dysplasia.";  
RL Gene 255:307-316(2000).  
DR EMBL: AF023169; AAC62178.2; -;  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR000885; Fib\_collagen\_C.  
DR InterPro: IPR001007; VWF\_C.  
DR Pfam: PF01410; COLFI; 1.  
DR Pfam: PF01391; Collagen; 18.  
DR Pfam: PF00093; VWC; 1.  
DR ProDom: PD000007; Collagen; 4.  
DR ProDom: PD002078; Fib\_collagen\_C; 1.  
DR SMART: SM00038; COLFI; 1.  
DR SMART: SM00214; VWC; 1.  
DR PROSITE: PS01208; VWF; 1.  
KM Collagen.  
SQ SEQUENCE 1487 AA; 141875 MW; 25873EAC1E311DB8 CRC64;

Query Match 62.5%; Score 40; DB 6; Length 1487;  
Best Local Similarity 87.5%; Pred. No. 82;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 POGIAGOR 9  
Db 973 POGIAGOR 980  
RESULT 19  
ID 09KM17 PRELIMINARY; PRT; 211 AA.  
AC 09KM17;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE DNA insertion sequence IS1421 (Fragment).  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF 301556;  
RA Hasebe A.;  
RT "Identification and nucleotide sequence of Ralstonia solanacearum  
RT insertion sequence IS1421, a new member of the IS427 subgroup of the  
RT ISS family.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB045354; BAA97978.1; -  
 DR InterPro: IPR002559; Transposase\_11.  
 DR Pfam: PF01609; Transposase\_11; 1.  
 FT NON\_TER  
 SQ SEQUENCE 211 AA; 23469 MW; 1AA5216B25677A8 CRC64;

Query Match  
 Best Local Similarity 77.8%; Score 39; DB 2; Length 211;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SPGINGOR 9  
 Db 13 SPAGVAGOR 21

RESULT 20  
 ID 09KKH7 PRELIMINARY; PRT; 373 AA.  
 AC 09KKH7;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE YSAW.  
 GN YSAW.  
 OS Versinia enterocolitica.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Versinia.  
 NCBI\_TaxID=630;  
 OX NCBI\_TaxID=630;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:20402121; PubMed:10931293;  
 RA Haller J.C., Carlson S., Pederson K.J., Pierson D.E.;  
 RT "A chromosomally-encoded type III secretion pathway in Versinia  
 RT enterocolitica is important in virulence.";  
 RL Mol. Microbiol. 36:1436-1446(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Carlson S., Pierson D.E.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Haller J.C., Carlson S., Pederson K.J., Pierson D.E.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF005744; AAF83328.1; -  
 DR InterPro: IPR003520; Invas\_inve.  
 DR Pfam: PF02523; Inve; 1.  
 SO SEQUENCE 373 AA; 41968 MW; 898E5979A9DF3ED CRC64;

Query Match  
 Best Local Similarity 59.4%; Score 38; DB 2; Length 373;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 GINGORNF 12  
 Db 287 GIRGOENFN 295

RESULT 21  
 ID 093NK8 PRELIMINARY; PRT; 373 AA.  
 AC 093NK8;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE YSAW.  
 GN YSAW.  
 OS Versinia enterocolitica.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Versinia.  
 NCBI\_TaxID=630;  
 OX NCBI\_TaxID=630;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-A127;  
 RA Foulter B.G.F., Mueller S., Purnelle B., Troisfontaines P.,  
 RA Cornuils G.R., (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF36954; AAK84112.1; -  
 DR InterPro: IPR003520; Invas\_inve.  
 DR Pfam: PF02523; Inve; 1.  
 SQ SEQUENCE 373 AA; 41968 MW; AFEF06F9863E904A CRC64;

Query Match  
 Best Local Similarity 59.4%; Score 38; DB 2; Length 373;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 GINGORNF 12  
 Db 287 GIRGOENFN 295

RESULT 22  
 ID 082804 PRELIMINARY; PRT; 695 AA.  
 AC 082804;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE At2g25930 protein (Hypothetical 77.2 kDa protein).  
 GN AT2G25930  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RX MEDLINE:20083487; PubMed:10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carreira A.O., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
 RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Town C.D., Kaul S.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004747; AAC31242.1; -  
 DR EMBL: AC005395; AAM15042.1; -  
 KW Hypothetical protein  
 SO SEQUENCE 695 AA; 77206 MW; 607A0720ED381C08 CRC64;

Query Match  
 Best Local Similarity 59.4%; Score 38; DB 10; Length 695;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 POGINGORNF 11  
 |||||:|:|:|

Db 608 POGISGSKSF 617

## RESULT 23

ID 004419 PRELIMINARY; PRT: 702 AA.

AC 004419; (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Nematode responsive protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-SHOOT

RX MEDLINE=20035823; PubMed=10571046;

RA Puzio P.S., Lausen J., Almeida-Engler J., Cai D., Gheysen G.,

RA Grundler G.M.W.;

RT "Isolation of a gene from Arabidopsis thaliana related to nematode

feeding structures."

RL Gene 239:163-172(1999).

DR EMBL, Y11994; CAA72719.1;

SQ SEQUENCE 702 AA; 78184 MW; F52F426657817954 CRC64;

## Query Match

Best Local Similarity 59.4%; Score 38; DB 10; Length 702;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 POGIAGORNF 11

Db 608 POGISGSKSF 617

## RESULT 24

ID 093485 PRELIMINARY; PRT: 809 AA.

AC 093485; (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Alpha 1 type I collagen (Fragment).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 CC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 NCBI\_TaxID=8022;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-FIBROBLAST

RA Saito M., Kunisaki N., Hiroto I., Aoki T., Ishida M., Urano N.,

RA Kimura S.;

RT "Partial characterization of cDNA clones encoding the three distinct

pro alpha chains of type I collagen from rainbow trout."

RL Fisheries Sci. 64:780-786(1998).

DR EMBL, AB008373; BAA3380.1;

DR InterPro: IPR000087; Collagen.

DR Pfam: PF01410; COLFI. 1.

DR Pfam: PF01391; Collagen. 9.

DR ProDom: PD000007; Collagen; 1.

DR ProDom: PD002078; Fib\_collagen\_C; 1.

DR SMART: SM00038; COLFI. 1.

DR KEGG collagen.

FT NON TER 1 1

SQ SEQUENCE 809 AA; 78164 MW; 68C056A7640FC81 CRC64;

## Query Match

Best Local Similarity 59.4%; Score 38; DB 13; Length 809;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 POGIAGOR 9

Db 296 POGIGCOR 303

## RESULT 25

ID 0910C0 PRELIMINARY; PRT: 1449 AA.

AC 0910C0; (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Collagen a1(I).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 CC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 NCBI\_TaxID=8022;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=21257802; PubMed=11358497;

RA Saito M., Takenouchi Y., Kunisaki N., Kimura S.;

RT "Complete primary structure of rainbow trout type I collagen

consisting of a1(I)a2(I)a3(I) heterotrimers."

RL Eur. J. Biochem. 268:2817-2827(2001).

DR EMBL: AB052835; BAB55661.1;

DR InterPro: IPR000087; Collagen.

DR InterPro: IPR001007; WF\_C.

DR Pfam: PF01410; COLFI. 1.

DR Pfam: PF01391; Collagen; 18.

DR Pfam: PF00093; WWC; 1.

DR ProDom: PD000007; Collagen; 2.

DR ProDom: PD002078; Fib\_collagen\_C; 1.

DR PROSITE: PS01208; WWC; UNKNOWN\_1.

SQ SEQUENCE 1449 AA; 137117 MW; 62EEF8A7BFD652B8 CRC64;

## Query Match

Best Local Similarity 87.5%; Score 38; DB 13; Length 1449;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 POGIAGOR 9

Db 936 POGIGCOR 943

## RESULT 26

ID 09A180 PRELIMINARY; PRT: 1647 AA.

AC 09A180; (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative cell envelope proteinase.  
 OS Streptococcus pyogenes.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 CC Streptococcaceae; Streptococcus.  
 NCBI\_TaxID=1314;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;

RX MEDLINE=21192684; PubMed=11296296;

RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.;

RA Primeaux C., Sezate S., Suvarov A.N., Kenton S., Lai H.S., Lin S.P.,

RA Qian Y., Jia H.G., Najat F.Z., Ren Q., Zhu H., Song L., White J.,

RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;

RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."

DR EMBL: AE006503; AAK33444.1;

DR HSP: P00782; 2SRT.



RESULT 29

097550 PRELIMINARY; PRT; 84 AA.

AC 097550: PRELIMINARY; PRT; 84 AA.

DT 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Hypothetical protein SP0563.

GN SP0563.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

NCBI\_TaxID=1313;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TIGR4;

RX MEDLINE=21357209; PubMed=11463916;

RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J., Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D., Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L., McDonald L.A., Feldblyum T.V., Angiuoli S., Dickson T., Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;

RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";

RL Science 293:498-506(2001);

DR EMBL: AE007366; AAK74719.1; -

DR TIGR: SP0563; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 84 AA; 10142 MW; 5F3784A94EBF1BCD CRC64;

Query Match 57.8%; Score 37; DB 16; Length 84;

Best Local Similarity 50.0%; Pred. No. 15;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 POGIAGORNF 11

Db 75 PEGIFGERNY 84

RESULT 30

095061 PRELIMINARY; PRT; 328 AA.

AC 095061;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Wasp interactor protein (Fragment).

GN WISP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

RA Wilson L.A., Fields D., Cruz L., Friesen J., Siminovitch K.A.;

RT "Wasp interactor protein (WISP): a novel SH3 protein that interacts with the Wiskott Aldrich Syndrome Protein (WASP)."

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

DR EMBL: AF001629; AAD00898.1; -

DR InterPro: IPR001683; PX.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00787; PX; 1.

DR Pfam: PF00018; SH3; 1.

DR PRINTS: PRO00452; SH3DOMAIN.

DR PRODOM: PD000066; SH3; 1.

DR SMART: SM00326; SH3; 1.

DR PROSITE: PS00002; SH3; 1.

KW SH3 domain.

FT NON\_TER 328 328

SQ SEQUENCE 328 AA; 35815 MW; 6195EF917E7485C7 CRC64;

Query Match 57.8%; Score 37; DB 4; Length 328;

Best Local Similarity 63.6%; Pred. No. 64;

Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 POGIAGORNF 12

Db 136 POGAGORNTN 146

RESULT 31

08704 PRELIMINARY; PRT; 463 AA.

AC 08704;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE LD03583p.

GN CG4023.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephyroidae; Drosophilidae; Drosophila.

NCBI\_TaxID=7227;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J., Champagne M., Chavez C., Dorsett V., Farfan D., Fise E., George R., Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M., Celinker S.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY069319; AAL39464.1; -

SQ SEQUENCE 463 AA; 53154 MW; 30F6F291542247E CRC64;

Query Match 57.8%; Score 37; DB 5; Length 463;

Best Local Similarity 54.5%; Pred. No. 92;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 SPOGIAGORNF 11

Db 435 NPDGIAGRSNF 445

RESULT 32

042434 PRELIMINARY; PRT; 496 AA.

AC 042434;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Cytokeratin 8 (Fragment).

GN K8.

OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae;

OC Notophthalmus.

NCBI\_TaxID=8316;

OX [1]

RP SEQUENCE FROM N.A.

RC TISSUE=REGENERATING LIMB BLASTEMA;

RX MEDLINE=98075904; PubMed=9415422;

RA Corcoran J.P., Ferretti P.;

RT "Keratin 8 and 18 expression in mesenchymal progenitor cells of differentiating limbs is associated with cell proliferation and differentiation."

RL Dev. Dyn. 210:355-370(1997).

CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

DR EMBL: AJ001296; CAA04656.1; -

DR InterPro: IPR001664; IF.

DR InterPro: IPR003054; Keratin\_II.  
 DR Pfam: PF00038; Filament; 1.  
 DR PRINTS: PR01276; TYPEKERATIN.  
 DR PROSITE: PS00226; IF: 1.  
 KW Colled coil; Intermediate filament; Keratin.  
 FT NON\_TER 1  
 SQ SEQUENCE 496 AA; 5512 MW; D327911FBB7A1CBA CRC64;

Query Match  
 Best Local Similarity 57.8%; Score 37; DB 13; Length 496;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 SPOGIAGORNF 12  
 DB 2 SARGAPCORNF 13

RESULT 33

OBXDP3 PRELIMINARY; PRT; 739 AA.  
 AC 08XDP3;  
 DT 01-MAR-2002 (TRENBLREL. 20, Created)  
 DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
 DE Biotin sulfoxide reductase.  
 GN BISC OR Z4976 OR ECS4436.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RA MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Grobbeck E.J., Hackert J., Klink S., Boutin A., Shao Y., Miller L., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RA MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsuno E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T., Kubura S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL: AE005582; BAG58700.1; -;  
 DR EMBL: AP002565; BAB37859.1; -;  
 DR InterPro: IPR001467; Prok\_Mboxred.  
 DR Pfam: PF00384; Molybdopterin; 1.  
 DR Pfam: PF01568; Molybdopterin; 1.  
 DR TIGRFAMS: TIGR00509; b1sc\_Fam; 1.  
 DR PROSITE: PS00490; MOLYBDOPTERIN\_PROK\_2; 1.  
 DR PROSITE: PS00932; MOLYBDOPTERIN\_PROK\_3; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 739 AA; 81961 MW; 23675207DD4FE6C3 CRC64;

Query Match  
 Best Local Similarity 57.8%; Score 37; DB 16; Length 739;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 SPOGIAGORNF 11  
 DB 34 NPOGICORNF 44

RESULT 34  
 O8U724 PRELIMINARY; PRT; 771 AA.

AC 08U724;  
 DT 01-JUN-2002 (TRENBLREL. 21, Created)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
 DE ATP-dependent DNA ligase.  
 GN ATU4637 OR AGR\_L\_502.  
 OS Agrobacterium tumefaciens (Strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Klatzma J.P., Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L., Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D., Sr., Chapman P., Clendenning J., Deatherage G., Gillet M., Grant C., Kutayyan T., Levy R., Li M.-J., McClelland E., Palmeri A., Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M., Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V., Neeter E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58";  
 RL Science 294:2317-2323(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle B., Gattling S., Miller N., Blanchard M., Ourollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Houmel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F., Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B., Ralagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G., Clato C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58";  
 RL Science 294:2323-2328(2001).  
 DR EMBL: AE009391; AAL45426.1; -;  
 DR EMBL: AE008224; AAK88815.1; -;  
 KW Ligase; Complete proteome.  
 SQ SEQUENCE 771 AA; 85604 MW; BE8B7F00DA84207B CRC64;

Query Match  
 Best Local Similarity 70.0%; Score 37; DB 16; Length 771;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 POGIAGORNF 11  
 DB 594 PEGIEGORNF 603

RESULT 35  
 O9V6K7 PRELIMINARY; PRT; 865 AA.  
 AC 09V6K7;  
 DT 01-MAY-2000 (TRENBLREL. 13, Created)  
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE CG4023 protein.  
 GN CG4023.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amandites P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champagne B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abri J.F., Abdayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Durbin K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garr N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strung R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003819; AAF58416.1; -  
 DR FLYBase: FBgn003801; CG4023.  
 DR InterPro: IPR000461; Alpha\_amyase.  
 DR InterPro: IPR004193; Isoamyase.  
 DR Pfam: PF00128; alpha-amyase; 1.  
 DR Pfam: PF02922; isoamyase; 1.  
 SQ SEQUENCE 865 AA; 99760 MW; 9B24CB9B9F01518A CRC64;

Query Match 57.8%; Score 37; DB 5; Length 865;  
 Best Local Similarity 54.5%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SPOGIAGORNF 11  
 Db 657 NPGYAGRSNF 667

RESULT 36  
 ID 093EC2 PRELIMINARY; PRT; 1158 AA.  
 AC 093EC2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Impl.  
 GN IMPL.  
 OS Rhizobium leguminosarum (biovar trifolii).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=386;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bladergroen M.R., Badelt K., Stronk O.P., Lugtenberg E.J.J.,  
 RA Spaik H.P.;  
 "An avirulence locus of a symbiotic Rhizobium leguminosarum strain is

RT involved in temperature-dependent protein secretion."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF361470; AAL17784.1; -  
 SQ SEQUENCE 1158 AA; 127575 MW; 3374EAFBA62B5C5 CRC64;

Query Match 57.8%; Score 37; DB 2; Length 1158;  
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 SPOGIAGORNF 12  
 Db 170 SVQIGGTRNCN 181

RESULT 37  
 ID 091717 PRELIMINARY; PRT; 1486 AA.  
 AC 091717;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Alpha-1 type II collagen.  
 GN COL2A1.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE:92011898; PubMed:1918153;  
 RA Su M.W., Suzuki H.R., Bleker J.J., Solursh M., Ramirez F.,  
 RT "Expression of two nonallelic type II procollagen genes during *Xenopus*  
 RT laevis embryogenesis is characterized by stage-specific production of  
 RT alternatively spliced transcripts."  
 RL J. Cell Biol. 115:565-575(1991).  
 DR EMBL: M63595; AAA49678.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF00391; Collagen; 18.  
 DR Pfam: PF00093; VWC; 1.  
 DR ProDom: PD000007; Collagen; 4.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00214; VWC; 1.  
 DR PROSITE: PS01208; VWF; 1.  
 KW Collagen.  
 SQ SEQUENCE 1486 AA; 142263 MW; 4AAA95772341042F CRC64;

Query Match 57.8%; Score 37; DB 13; Length 1486;  
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 POGIAGOR 9  
 Db 975 POGISGR 982

RESULT 38  
 ID 045769 PRELIMINARY; PRT; 369 AA.  
 AC 045769;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE T10C6.7 protein.  
 GN T10C6.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;



```

RN [1]
RP SEQUENCE FROM N.A.
RX MCMurray A.A.;
RN Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99069613; PubMed-9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
  investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL:293388; CAB07663.1;
DR InterPro: IPR001810; F-box;
DR Pfam: PF00646; F-box; 1.
DR PROSITE: PS50181; FBOX; 1.
SQ SEQUENCE 369 AA; 43506 MW; 00DEC66EAF291041 CRC64;

Query Match
Best Local Similarity 57.0%; Score 36.5; DB 5; Length 369;
Matches 9; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

OY 1 SPOGIAGORNFN 12
Db 331 SPOGIAGORN-FN 341

RESULT 39
ID 09Y3P3 PRELIMINARY; PRT; 100 AA.
AC 09Y3P3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE BK984G1.1 (Supported by FGENSES) (Fragment).
GN BK984G1.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matchews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL031186; CAB41538.1;
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 1.
FT NON_TER 1
SQ SEQUENCE 100 AA; 10473 MW; 220ED333CCA05668 CRC64;

Query Match
Best Local Similarity 56.2%; Score 36; DB 4; Length 100;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 POGIACOR 9
Db 8 POGIACOR 15

RESULT 40
ID 0915P4 PRELIMINARY; PRT; 149 AA.
AC 0915P4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Probable type II secretion system protein.
GN PA0681.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxId=287;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Phan X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brickman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Gardner R.L., Goltzy L., Tolentino E., Westbrook-Wedman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Loy S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
  opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS (BY SIMILARITY).
CC -1- SUBUNIT: THE PILI ARE POLAR FLEXIBLE FILAMENTS OF ABOUT 5.4
CC NANOMETERS DIAMETER AND 2.5 MICROMETERS AVERAGE LENGTH; THEY
CC CONSIST OF ONLY A SINGLE POLYPEPTIDE CHAIN ARRANGED IN A HELICAL
CC CONFIGURATION OF FIVE SUBUNITS PER TURN IN THE ASSEMBLED PILUS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE N-ME-PHE PILI FAMILY.
CC -1- SIMILARITY: BELONGS TO THE PULG/OUTG/XPSG/EXES/XCPT FAMILY.
DR EMBL: AE004503; AAC04070.1;
DR InterPro: IPR000983; Bac_GSPG.
DR InterPro: IPR002416; Bac_GSPH.
DR InterPro: IPR001120; Prok_N_methyltn.
DR PRINTS: PR00813; BACTERIALGSPG.
DR PRINTS: PR00885; BACTERIALGSPH.
DR PROSITE: PS00409; PROKAR_NTER_METHYL. 1
KW Fimbria; Methylation; Transport; Complete proteome.
SQ SEQUENCE 149 AA; 16219 MW; 0D3E6315BC63BC28 CRC64;

```

```

OY 1 SPOGIAGORNF 11
Db 9 SPOGIAGORNF 19

Query Match
Best Local Similarity 56.2%; Score 36; DB 16; Length 149;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 SPOGIAGORNF 11
Db 9 SPOGIAGORNF 19

RESULT 41
ID 0920V4 PRELIMINARY; PRT; 409 AA.
AC 0920V4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE UDP-Gal:betaGalNAc beta 1,3-galactosyltransferase II (Fragment).
GN B3GT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CAST/EI;
RA Liu Y., Kitano T., Koide T., Shirosaki T., Moriwaki K., Saitou N.;
RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
  Five Mus musculus subspecies.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB039147; BAB86871.1;
DR InterPro: IPR002659; Galactosyl_T.
DR Pfam: PF01762; Galactosyl_T; 1.
DR GlycoSyltransferase; Transferase.
FT NON_TER 1
SQ SEQUENCE 409 AA; 47247 MW; 2423654F23A8F0E3 CRC64;

Query Match
Best Local Similarity 56.2%; Score 36; DB 11; Length 409;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SPOGIAGORN 10

```

Db 99 SPOGVTGLON 108

## RESULT 42

Q920V3 PRELIMINARY; PRT; 409 AA.  
 AC Q920V3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).  
 GN B3GT2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HMI/MSF;  
 RA Liu Y., Kitano T., Koide T., Shirosaki T., Moriwaki K., Saitou N.;  
 RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of  
 RT Five Mus musculus subspecies."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB039151; BAB68675.1; -;  
 DR InterPro: IPR002659; Galactosyl\_T.  
 DR Pfam: PF01762; Galactosyl\_T.1.  
 KW Glycosyltransferase; Transferase.  
 FT NON\_TER 1  
 FT NON\_TER 409  
 SQ SEQUENCE 409 AA; 47220 MW; 2D2BEC4D68E8463D CRC64;

## Query Match

Best Local Similarity 56.2%; Score 36; DB 11; Length 409;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SPOGIAGORN 10  
 ||||: 1:1  
 Db 99 SPOGVTGLON 108

## RESULT 43

Q920V2 PRELIMINARY; PRT; 409 AA.  
 AC Q920V2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).  
 GN B3GT2.  
 OS Mus spicilegus (Steppe mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ZBN;  
 RA Liu Y., Kitano T., Koide T., Shirosaki T., Moriwaki K., Saitou N.;  
 RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of  
 RT Five Mus musculus subspecies."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB039153; BAB68677.1; -;  
 DR InterPro: IPR002659; Galactosyl\_T.  
 DR Pfam: PF01762; Galactosyl\_T.1.  
 KW Glycosyltransferase; Transferase.  
 FT NON\_TER 1  
 FT NON\_TER 409  
 SQ SEQUENCE 409 AA; 47383 MW; 0DC51AB6335A58DE CRC64;

## Query Match

Best Local Similarity 56.2%; Score 36; DB 11; Length 409;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SPOGIAGORN 10

Db 99 SPOGVTGLON 108

## RESULT 44

Q91V58 PRELIMINARY; PRT; 409 AA.  
 AC Q91V58;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).  
 GN B3GT2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BLG/MSF, AND NUL/MSF;  
 RA Liu Y., Kitano T., Koide T., Shirosaki T., Moriwaki K., Saitou N.;  
 RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of  
 RT Five Mus musculus subspecies."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB039146; BAB68670.1; -;  
 DR EMBL: AB039152; BAB68676.1; -;  
 DR InterPro: IPR002659; Galactosyl\_T.  
 DR Pfam: PF01762; Galactosyl\_T.1.  
 KW Glycosyltransferase; Transferase.  
 FT NON\_TER 1  
 FT NON\_TER 409  
 SQ SEQUENCE 409 AA; 47196 MW; 06D262B08B090F4 CRC64;

## Query Match

Best Local Similarity 56.2%; Score 36; DB 11; Length 409;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SPOGIAGORN 10  
 ||||: 1:1  
 Db 99 SPOGVTGLON 108

## RESULT 45

Q91V58 PRELIMINARY; PRT; 409 AA.  
 AC Q91V58;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).  
 GN B3GT2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/10SNJ, BFW/2MSF, AND PGN2;  
 RA Liu Y., Kitano T., Koide T., Shirosaki T., Moriwaki K., Saitou N.;  
 RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of  
 RT Five Mus musculus subspecies."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB039144; BAB68668.1; -;  
 DR EMBL: AB039145; BAB68669.1; -;  
 DR EMBL: AB039149; BAB68673.1; -;  
 DR InterPro: IPR002659; Galactosyl\_T.  
 DR Pfam: PF01762; Galactosyl\_T.1.  
 KW Glycosyltransferase; Transferase.  
 FT NON\_TER 1  
 FT NON\_TER 409  
 SQ SEQUENCE 409 AA; 47233 MW; 9DBE67873273F13A CRC64;

## Query Match

Best Local Similarity 56.2%; Score 36; DB 11; Length 409;

Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 SPOGAGORN 10  
1111:1:1  
DB 99 SPOGVTGLON 108

## RESULT 46

QY 091V19 PRELIMINARY; PRT; 409 AA.  
AC 091V19  
DB 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE UDP-Gal:betaglcNAC beta 1,3-galactosyltransferase II (Fragment).  
GN B3GT2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSM/MSF, AND SWN/MSF;  
RA Liu Y., Kitano T., Koide T., Shirotschi T., Moriwa K., Salou N.;  
RT "Conspicuous differences among Gene Genealogies of 21 Nuclear Genes of  
RT Five Mus musculus subspecies."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB039148; BAB6672.1; -;  
DR EMBL; AB039150; BAB6674.1; -;  
DR InterPro: IPR002659; Galactosyl-T.  
DR Pfam: PF01762; Galactosyl-T.  
KW Glycosyltransferase; transferase.  
FT NON\_TER 1  
FT 409  
SQ SEQUENCE 409 AA; 47206 MW; D7BD514EE40A0157 CRC64;

Query Match 56.2%; Score 36; DB 11; Length 409;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SPOGAGORN 10  
1111:1:1  
DB 99 SPOGVTGLON 108

## RESULT 47

QY 054905 PRELIMINARY; PRT; 422 AA.  
AC 054905  
DB 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE UDP-galactose:beta-N-acetyl-glucosamine-beta-1,3-galactosyltransferase  
DE II (EC 2.4.1.-) (UDP-GAL:betaGLCNAC beta 1,3-GALACTOSYLTRANFERASE-II)  
DE (BETA3GALT-II).  
GN B3GT2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
RC STRAIN=129/SVJ;  
RA Henne T., Dinter A., Kuhner P., Mattu T.S., Rudd P.M., Berger E.G.;  
RT "Genomic cloning and expression of three murine UDP-galactose: beta-N-  
RT acetylglucosamine beta1,3-galactosyltransferase genes."  
RL J. Biol. Chem. 273:58-65(1998).  
CC -1- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF  
CC COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS  
CC WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT-I, GALT-II  
CC AND GALT-III ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH

CC DIFFERENT AFFINITY.  
CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETYL-BETA-D-GLUCOSAMINYL-  
CC GLYCOPETIDE -> UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-  
CC GLUCOSAMINYLGlyCOPEPTIDE.  
CC -1- COFACTOR: MANGANESE.  
CC -1- PATHWAY: GLYCOSYLATION.  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.  
CC IN OVARY, COLON AND LYMPH NODES. NO EXPRESSED IN TESTIS, UTERUS,  
CC STOMACH, KIDNEY, LIVER, LUNG, BONE MARROW, THYMUS AND SPLEEN.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-14 IS THE INITIATOR.  
DR EMBL; MG11349461; B391C2.  
DR InterPro: IPR002659; Galactosyl-T.  
DR Pfam: PF01762; Galactosyl-T.  
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Glycoprotein; Golgi stack; Manganese  
FT DOMAIN 1 24  
FT TRANSMEM 25 43  
FT DOMAIN 44 422  
FT CARBOHYD 75 75  
FT CARBOHYD 98 98  
FT CARBOHYD 119 119  
FT CARBOHYD 176 176  
FT CARBOHYD 226 226  
SQ SEQUENCE 422 AA; 49107 MW; D03D18A7484DCB03 CRC64;

## Query Match

Best Local Similarity 56.2%; Score 36; DB 11; Length 422;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SPOGAGORN 10  
1111:1:1  
DB 105 SPOGVTGLON 114

RESULT 48  
QY 09P8N1 PRELIMINARY; PRT; 436 AA.  
AC 09P8N1  
DB 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Cellulohydrolase.  
DE Coriolus versicolor.  
OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
OC Aphyllophorales; Coriolus.  
OX NCBI\_TaxID=57466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Novo C., Clemente A., Simoes F., Mendonca D., Matos J.;  
RT "Coriolus versicolor cellulohydrolase-like cDNA sequence."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF233583; AAF35251.1; -;  
DR HSSP: P00725; ZCBH.  
DR InterPro: IPR000254; CBD\_Fungal.  
DR InterPro: IPR001524; GH\_6.  
DR Pfam: PF00734; CBM\_1; 1.  
DR Pfam: PF01341; Glyco\_hydro\_6; 1.  
DR PRINTS: PR00733; GLHYDRLASE6.  
DR ProDom: PD001821; CBD\_Fungal; 1.  
DR ProDom: PD001821; CBD\_Fungal; 1.  
DR SMART: SM00236; fcbd; 1.  
DR PROSITE: PS00562; CBD\_FUNGAL; 1.  
DR PROSITE: PS00655; GLYCOSYL\_HYDROL\_F6\_1; 1.  
DR PROSITE: PS00656; GLYCOSYL\_HYDROL\_F6\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 436 AA; 46170 MW; E0ECC5AFBCA324D4 CRC64;

Query Match 56.2%; Score 36; DB 3; Length 436;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 OGAGORNF 11  
 11111111  
 DB 342 OGAGORNF 350

## RESULT 49

O96A84 PRELIMINARY; PRT; 441 AA.  
 ID O96A84  
 AC O96A84;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Putative emul protein precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Leimister C., Gessler M.;  
 RT "Expression of Emu genes during mouse embryogenesis."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ416090; CAC94777.1;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 2.  
 KW Signal.  
 FT SIGNAL.  
 SQ SEQUENCE 441 AA; 45292 MW; 676D491C4083E18E CRC64;

Query Match 56.2%; Score 36; DB 4; Length 441;  
 Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 POGIAGOR 9  
 11111111  
 DB 349 POGIAGOR 356

RESULT 50  
 O9KMS5 PRELIMINARY; PRT; 491 AA;  
 ID O9KMS5  
 AC O9KMS5;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Amc.  
 GN AMNC.  
 OS Pseudomonas sp.  
 OC Bacteria; Proteobacteria.  
 OX NCBI\_TaxID=306;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-AP-3;  
 RA Takenaka S., Murakami S., Kim Y.J., Aoki K.;  
 RT "Complete nucleotide sequence and functional analysis of the genes for  
 RT 2-aminophenol metabolism from Pseudomonas sp. AP-3."  
 RL Arch. Microbiol. 174:265-272(2001).  
 RN  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-AP-3;  
 RA Takenaka S., Murakami S., Shinke R., Hatakeyama K., Yukawa H.,  
 RT "Novel genes encoding 2-aminophenol 1,6-dioxygenase from Pseudomonas  
 RT sp. AP-3 growing on 2-aminophenol and catalytic properties of the  
 RT purified enzyme."  
 RL J. Biol. Chem. 272:14727-14732(1998).  
 CC - SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
 DR EMBL: AB020521; BAB03533.1;  
 DR HSSP: P56533; 1A4S.  
 DR InterPro: IPR002086; Aldehyde\_dehydr.  
 DR Pfam: PF00171; aldehyd. 1.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
 DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; UNKNOWN\_1.

KW Oxidoreductase.  
 SQ SEQUENCE 491 AA; 53722 MW; 1B196D37FF534551 CRC64;

Query Match 56.2%; Score 36; DB 2; Length 491;  
 Best Local Similarity 70.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 POGIAGORNF 11  
 11111111  
 DB 103 POGIAGORNF 112

Search completed: May 16, 2003, 10:41:03  
 Job time : 50 secs